CLASS-SPECIFIC POISSON DENOISING BY PATCH-BASED IMPORTANCE SAMPLING

Milad Niknejad, José M. Bioucas-Dias, Mário A. T. Figueiredo

Instituto de Telecomunicacoes
Instituto Superior Tecnico
Universidade de Lisboa, Portugal

ABSTRACT

In this paper, we address the problem of recovering images degraded by Poisson noise, where the image is known to belong to a specific class. In the proposed method, a dataset of clean patches from images of the class of interest is clustered using multivariate Gaussian distributions. In order to recover the noisy image, each noisy patch is assigned to one of these distributions, and the corresponding minimum mean squared error (MMSE) estimate is obtained. We propose to use a self-normalized importance sampling approach, which is a method of the Monte-Carlo family, for the both determining the most likely distribution and approximating the MMSE estimate of the clean patch. Experimental results shows that our proposed method outperforms other methods for Poisson denoising at a low SNR regime.

Index Terms— Image denoising, Poisson noise, class-specific dataset, importance sampling.

1. INTRODUCTION

Recovering noisy images is a fundamental problem in image processing and computer vision. In many image denoising applications, especially in low intensity (photon-limited) images, the noise follows a Poisson distribution. Photon-limited acquisition scenarios are often found in astronomical and medical imaging, and in many other areas.

The classical formulation for Poisson denoising considers the task of recovering an underlying image, the pixels of which are stacked in a vector $\mathbf{x} \in \mathbb{R}_+^N$, from Poissonian observations $\mathbf{y} \in \mathbb{N}_0^N$, *i.e.*

$$\mathbb{P}(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^{N} \frac{e^{-\mathbf{x}_{[j]}} \mathbf{x}_{[j]}^{\mathbf{y}_{[j]}}}{\mathbf{y}_{[j]}!}. \tag{1}$$

A common approach to Poisson densoising proceeds as follows: first, a *variance stabilization transform* (VST) is applied (*e.g.*, the Anscombe transform [1]), which approximately turns the Poisson distribution into a Gaussian of unit variance. Then, one of the many existing methods for Gaussian denoising is applied to the transformed image. Finally an inverse transformation yields the image estimate [2, 3]. However, this approach is inaccurate for low intensity (low SNR) images, thus many methods denoise the Poissonian images without any transformation [4]. Our method belongs to this second category.

In many applications, the image to be recovered is known to belong to a certain class, such as text, face, or fingerprints. Knowing this information can help to learn a statistical prior that is better adapted to that specific class than a generic prior, and this prior can be learned from a dataset of clean images from that class. Recently this class-specific approach has been addressed in [5, 6], for denoising and deblurring under Gaussian noise.

Some patch-based methods for recovering images degraded by Gaussian noises use multivariate priors for image restoration to obtain *maximum a posteriori* (MAP) or *minimum mean squared error* (MMSE) estimates of the clean image. Most of these methods fit a mixture multivariate Gaussian distributions to patches, either from an external dataset [7], or from the noisy image itself [8, 9]. Although for Poisson denoising, patch-based methods have been proposed [4, 10], there is a lack of methods that use an explicit prior on image patches, due to the non-Gaussian forward model, which makes it difficult to compute the corresponding MAP or MMSE estimates. On the other hand, the fitted Gaussian mixture prior is an approximation of the true distribution that characterizes the image patches.

In this paper, we propose a class-adapted method to recover images degraded by Poisson noise using a dataset of clean images of the same class. In our method, a mixture of multivariate Gaussian distribution is fitted to the external dataset of clean image patches. For each noisy patch of the observed image, we use the self-normalized importance sampling [11], which is an approach from the family of Monte-Carlo methods, to both choose the proper prior distribution and approximate the MMSE estimation of each patch. Using importance sampling allows us to approximate the MMSE estimate using the true distribution of image patches, if the sam-

ples are the patches of clean images in the dataset. Furthermore, our proposed method can be extended to other forward models with known noise distribution, and different prior densities of clean patches.

In the following sections, we first briefly review the self-normalizing importance sampling approach. Then, we describe the proposed method for MMSE Poisson image denoising based on self-normalizing importance sampling. Section 4 reports experimental results.

2. SELF-NORMALIZED IMPORTANCE SAMPLING

Our approach uses a technique from the family of Monte-Carlo methods, called *self-normalized importance sampling* (SNIS) [11, 12], which we will now briefly review. Assume that the objective is to compute (or approximate) $\mathbb{E}[f(\mathbf{X})]$, the expectation of a function $f : \mathbb{R}^d \to \mathbb{R}^p$ of a random variable $X \in \mathbb{R}^d$. Denoting the support of the random variable $\mathbf{X}$ as $\mathcal{R} \subset \mathbb{R}^d$, the expectation is given by

$$\mathbb{E}[f(\mathbf{x})] = \int_{\mathcal{R}} f(\mathbf{x})p(\mathbf{x})d\mathbf{x}. \tag{2}$$

Consider also that only an un-normalized version $\tilde{p}(\mathbf{x}) = c\,p(\mathbf{x})$ of the probability density function $p(\mathbf{x})$ of $\mathbf{X}$ is known, where c is unknown. Let $\tilde{q}(\mathbf{x}) = d\,q(\mathbf{x})$ be another un-normalized density, where the normalizing constant d is also unknown, but from which samples can be more efficiently obtained than from $p(\mathbf{x})$.

SNIS produces an estimate of $\mathbb{E}[f(\mathbf{X})]$ given by

$$\hat{\mathbb{E}}_n[f(\mathbf{x})] = \frac{\sum\limits_{j=1}^{n} f(\mathbf{x}_j)w(\mathbf{x}_j)}{\sum\limits_{j=1}^{n} w(\mathbf{x}_j)}, \tag{3}$$

where $w(\mathbf{x}_j) = \frac{\tilde{p}(\mathbf{x}_j)}{\tilde{q}(\mathbf{x}_j)}$, and $\mathbf{x}_1, ..., \mathbf{x}_n$ are n independent and identically distributed samples from distribution $q(\mathbf{x})$ [11, 12]. This approximation can be shown to converges to the true value of $\mathbb{E}[f(\mathbf{x})]$ as n goes to infinity.

3. PROPOSED METHOD

3.1. Prior Learning

The first step of the proposed method is to fit a set of K multivariate Gaussians to the patches in the external dataset of clean images. We adopt the so-called *classification EM* (CEM) algorithm [13], rather than a standard EM algorithm as in [8, 6, 7], for reasons explained below. The CEM algorithm works by alternating between the following two steps (after being initialized by standard K-means clustering):

1. From the set of patches assigned to each cluster $k \in \{1, ..., K\}$, denoted $\mathbf{X}_k$, obtain estimates of the mean μ_k and covariance matrix Σ_k of the corresponding Gaussian density, which are simply the sample mean and the sample covariance of the patches in $\mathbf{X}_k$.

2. Assign each patch to the cluster under which it has the highest likelihood, that is, patch $\mathbf{x}_j$ is assigned to $\mathbf{X}_k$ if

$$k = \arg \max_m \mathcal{N}(\mathbf{x}_j; \mu_m, \Sigma_m),$$

where, as usual, $\mathcal{N}(\mathbf{x}; \mu, \Sigma)$ denotes a Gaussian density of mean μ and covariance Σ, computed at $\mathbf{x}$.

In the denoising step, each noisy patch will be assigned to one of these clusters. The reason for this kind of clustering is that in the simple importance sampling approach, for a fixed number of samples n, the MSE of the estimator is proportional to the variance of samples being averaged [12]. Similarly in the multivariate case, for a fixed n, it has been shown that the MSE of the estimator of a particular entry in the vector decreases as the variance of the samples of that entry, given the noisy patch, decreases [14]. Consequently, since in practice we use a limited number of patches from the external dataset, by clustering this way we expect to reduce the estimator variance, without increasing too much the number of samples.

It should be noted that the above procedure needs to be applied once for a given dataset of class-specific images.

3.2. Image denoising

In the denoising step, each patch is assigned to one of the clusters obtained in the learning stage. However, we only have noisy patches, thus the assignment is not trivial. If the noise was Gaussian, the assignments could be made in closed-form, but this is not the case with Poisson observations. Our main contribution is a new method, based on SNIS, to simultaneously determine the cluster and estimate the clean patch.

Define the random variable $k_i \in \{1, ..., K\}$, which indicates the cluster to which the i-th patch belongs, and denote the corresponding distribution as $p(\mathbf{x}_i|k_i)$. Having a set of learned cluster distributions, our objective is to solve the following simultaneous classification and MMSE estimation problem, given a noisy patch $\mathbf{y}_i$:

$$(\hat{\mathbf{x}}_i, \hat{k}_i) = \arg \min_{(\mathbf{u}, k)} \int_{\mathbb{R}_+^m} \|\mathbf{u} - \mathbf{x}\|_2^2 \; p(\mathbf{x}|\mathbf{y}_i, k) \; d\mathbf{x}, \tag{4}$$

where m is the number of pixels in each patch. In other words, we seek the estimate and the cluster that yield the minimum MSE. As shown below, we solve the above problem by the alternating minimization approach, but first we need to address the problem of how to approximate this integral.

First, using Bayes rule and the fact that $p(\mathbf{y}_i|\mathbf{x}, k) = p(\mathbf{y}_i|\mathbf{x})$ (*i.e.*, given the clean patch, the noisy one does not

depend on the cluster), the integral in (4) can be written as

$$\mathbb{E}[\|\mathbf{x}-\mathbf{u}\|_2^2|\mathbf{y}_i,k] = \int_{\mathbb{R}_+^m} \|\mathbf{u}-\mathbf{x}\|_2^2 \, \frac{p(\mathbf{y}_i|\mathbf{x})\,p(\mathbf{x}|k)}{p(\mathbf{y}_i|k)} \, d\mathbf{x}. \quad (5)$$

Then, using SNIS, the above integral can be approximated by

$$\hat{\mathbb{E}}_n[\|\mathbf{x}-\mathbf{u}\|_2^2|\mathbf{y}_i,k] = \frac{\sum_{j=1}^n \|\mathbf{u}-\mathbf{x}_j\|_2^2 \, p(\mathbf{y}_i|\mathbf{x}_j)}{\sum_{j=1}^n p(\mathbf{y}_i|\mathbf{x}_j)} \quad (6)$$

where the $\mathbf{x}_j$, for $j = 1,...,n$ are samples from the distribution $p(\mathbf{x}|k)$. Exploiting the self-normalized importance sampling formula (3), in (6) we considered $\tilde{p}(\mathbf{x}) = p(\mathbf{y}_i|\mathbf{x})\,p(\mathbf{x}|k)$, with the unknown constant $c = 1/p(\mathbf{y}_i|k)$ and $q(\mathbf{x}) = p(\mathbf{x}|k)$.

The minimization with respect to k in (4), while $\mathbf{u}$ is fixed to be $\hat{\mathbf{x}}_i$, can then be approximated as

$$\hat{k}_i = \arg\min_k \hat{\mathbb{E}}_{n_2}[\|\mathbf{x}-\hat{\mathbf{x}}_i\|_2^2|\mathbf{y}_i,k]. \quad (7)$$

Note that for computing (7), n_2 samples from each distribution $p(\mathbf{x}|k)$ are randomly extracted.

The next step towards an alternating minimization method is a way to minimize $\mathbb{E}_n[\|\mathbf{x}-\mathbf{u}\|_2^2|\mathbf{y}_i,k]$ with respect to $\mathbf{u}$, for a given $k = \hat{k}_i$. This is the well-known MMSE estimation criterion, which is given by the posterior expectation $\mathbb{E}(\mathbf{x}|\mathbf{y}_i,\hat{k}_i)$. Computing this expectation cannot be done in closed-form, under the Poisson observation model and Gaussian prior $p(\mathbf{x}|k) = \mathcal{N}(\mathbf{x};\mu_k,\Sigma_k)$. To tackle this difficulty, we resort again to SNIS,

$$\hat{\mathbf{x}}_i = \hat{\mathbb{E}}_{n_1}[\mathbf{x}|\mathbf{y}_i,\hat{k}_i] = \frac{\sum_{j=1}^{n_1} \mathbf{x}_j p(\mathbf{y}_i|\mathbf{x}_j)}{\sum_{j=1}^{n_1} p(\mathbf{y}_i|\mathbf{x}_j)} \quad (8)$$

where $\mathbf{x}_1,...,\mathbf{x}_{n_1}$ are samples drawn from the distribution $p(\mathbf{x}|\hat{k})$. The approximation in (8) has been used before for patch-based denoising [15], but without noticing its connection to SNIS.

Since the multivariate Gaussian distributions fitted to the set of patches is only a crude approximation of the true distribution of the patches, samples from $p(\mathbf{x}|k)$ may yield poor results in the SNIS approximation. Instead of sampling from $p(\mathbf{x}|k)$, we directly sample form the set patches in the dataset assigned to that k-th cluster, $\mathbf{X}_k$, which turns out to yield better results.

We initialize our algorithm by $\mathbf{u} = \mathbf{y}_i$. Our algorithm then alternates between (8) and (7), to assign each patch to a cluster and obtain an MMSE estimate of the patch. The procedure of iterative clustering and denoising has been used

in some well-known patch-based denoising methods [16, 7]. The patches are returned to the original position in the image and are averaged in the overlapping pixels.

Note that our method can be applied to any noise model with a known distribution. In this paper, we only consider Poissonian noise and we leave other possible noise distributions for future work. For the Poisson case, the conditional distribution of a noisy patch $\mathbf{y}$, given a clean patch $\mathbf{x} \in \mathbb{R}_+^m$, is as shown in (1), with $N = m$.

- Initialization: cluster the training patches into K clusters, $\mathbf{X}_1^{(1)}\ldots\mathbf{X}_K^{(1)}$, via the k-means algorithm

- Main loop: for $i = 1,\ldots,I$

  - Estimate the parameter of Gaussian distributions $\Theta_k^{(i)} = \{\boldsymbol{\mu}_k^{(i)},\boldsymbol{\Sigma}_k^{(i)}\}$ as the sample mean and the sample covariance matrix of $\mathbf{X}_k^{(i)}$

  - For each patch $\mathbf{x}_j$, assign it to a cluster, *i.e.*, put $\mathbf{x}_j$ into $\mathbf{X}_{\hat{k}_j}^{(i)}$, where

    $$\hat{k}_j = \arg\max_k \mathcal{N}(\mathbf{x}_j;\boldsymbol{\mu}_k^{(i)},\boldsymbol{\Sigma}_k^{(i)})$$

- Output: $\{\mathbf{X}_1^{(I)}\ldots\mathbf{X}_K^{(I)}\}$

Fig. 1: Learning the prior from a class-specific dataset.

4. PRACTICAL CONSIDERATIONS AND EXPERIMENTAL RESULTS

The reported results in this section were obtained on the Gore dataset of face images [17] and on the dataset of text images used in [5]. For each dataset, 5 images are randomly chosen as test images and the rest are chosen for training. For the face images, we extracted 95×10^3 patches from the training data, whereas from the text dataset, 75×10^3 patches were extracted.

The described method is computationally expensive, if it is applied to all patches in the dataset. However, the key to reduce the computational complexity is to limit the number of patch samples used for estimating the clean patches and determining the cluster i.e. n_1 and n_2. Our results in this section show that for a very limited number of samples, we obtain acceptable results, which outperforms other Poisson denoising methods for the tested datasets. For determining the cluster, we set $n_2 = 30$ which is overall 600 patches for all $k = 20$ clusters and is less than 1% of the samples in each external datasets. The number of samples, n_1, used for denoising each patch was set to 300. So, overall for each noisy patch we

Table 1: Denoising PSNR results (in dB) for different peak values of face images in the Gore face database [17]. The results are averaged over 5 test images.

	2	5	10	15
NL-PCA [4]	19.69	22.87	23.80	25.01
VST+BM3D [3]	20.80	23.15	24.79	25.41
Poisson NL-means [18]	21.12	23.41	24.73	25.32
P4IP [19]	20.03	23.78	24.88	25.84
Our method	**21.31**	**23.95**	**25.78**	**27.40**

processed 900 patches which is in computational complexity roughly similar to an internal non-local denoising with the patches constrained in 30×30 window. Unlike the original non-local means, which only the central pixel of each patch is denoised, in our method the whole patch is denoised by (8), the patches are then returned to the original position in the image and are averaged in the overlapping pixels. In order to further reduce the computational complexity, we extract the patches from the noisy image every 2 pixels along the row and the column of the image.

In Table 1, the average result of PSNR of the 5 tested images for the face dataset are compared for different methods. The results of our method result from two iterations of the alternating minimization approach. We found that increasing the number of iteration to more than two does not noticeably increase the quality of obtained image estimate, while it obviously increases the computational cost. The reason may lie in the fact that the discrete variables k_i computed in (7) stabilize after a couple of iterations.

In Fig. (2), an example of denoising results for the face images with the peak value of 10 using the dataset in [17] is illustrated. It can be seen that our method also improves noticeably the visuals quality of image.

An example of denoising text images is shown in Fig. (3). It can be seen that the general image denoising methods fail to reconstruct the true image properly. However our class-specific method outperforms by a relatively high margin.

5. CONCLUSION

We proposed a method for class-specific denoising of images degraded by Poisson noise. We proposed to use a method from the Monte-Carlo family, called self-normalized importance sampling, in order to determine the prior and approximate the MMSE estimation. The sampling approach allowed us to approximate the MMSE using the true underlying priors rather than fitted multivariate Gaussian distributions. Our results showed that our method outperforms other methods in both PSNR and visual quality.
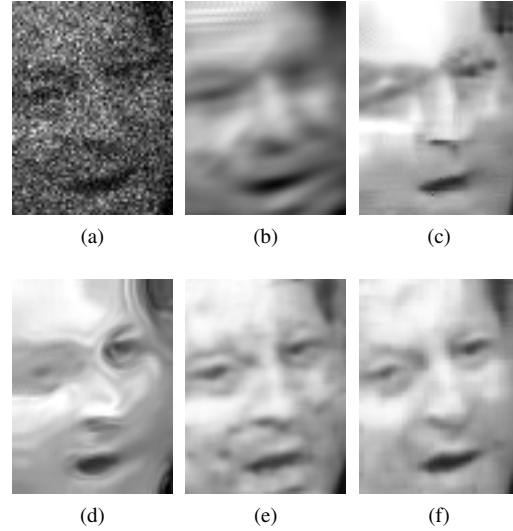
Fig. 2: An example of denoising of a face image using the Gore dataset ; (a) Noisy image (Peak value=10); (b) Non-local PCA (PSNR=22.60); (c) VST+BM3D (PSNR=24.79); (d) Poisson non-local means (PSNR=24.55); (e) Proposed (First iteration) (PSNR=25.88); (f) Proposed (Second iteration) (PSNR=26.40).
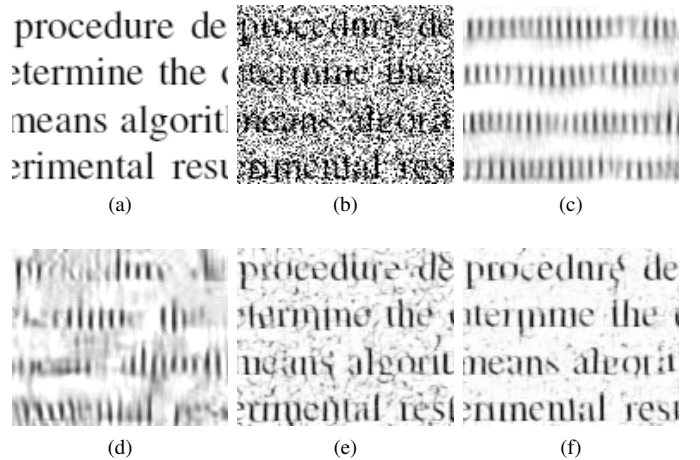
Fig. 3: Comparison of denoising of general methods with our class-specific one for recovering the text images; (a) Original image (b) Noisy image (Peak value=2) (c) Non-local PCA (PSNR=14.95) (d) VST+BM3D (PSNR=14.55); (e) Proposed (First iteration) (PSNR=17.21); (f) Proposed (Second iteration) (PSNR=18.64).

6. REFERENCES

[1] F. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948.

[2] F. Dupé, J. Fadili, and J. Starck, "A proximal iteration for deconvolving poisson noisy images using sparse representations," *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 310–321, 2009.

[3] M. Makitalo and A. Foi, "Optimal inversion of the Anscombe transformation in low-count Poisson image denoising," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 99–109, 2011.

[4] J. Salmon, Z. Harmany, C. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *Journal of Mathematical Imaging and Vision*, vol. 48, no. 2, pp. 279–294, 2014.

[5] E. Luo, S. Chan, and T. Nguyen, "Adaptive image denoising by targeted databases," *IEEE Transactions on Image Processing*, vol. 24, no. 7, pp. 2167–2181, 2015.

[6] A. Teodoro, J. Bioucas-Dias, and M. Figueiredo, "Image restoration and reconstruction using variable splitting and class-adapted image priors," in *IEEE International Conference on Image Processing*, 2016.

[7] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 479–486.

[8] A. Teodoro, M. Almeida, and M. Figueiredo, "Single-frame image denoising and inpainting using Gaussian mixtures," in *4th International Conference on Pattern Recognition Applications and Methods*, 2015.

[9] M. Niknejad, H. Rabbani, and M. Babaie-Zadeh, "Image restoration using Gaussian mixture models with spatially constrained patch clustering," *IEEE Transactions on Image Processing*, vol. 24, pp. 3624–3636, 2015.

[10] R. Giryes and M. Elad, "Sparsity-based Poisson denoising with dictionary learning," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5057–5069, 2014.

[11] T. Hesterberg, "Weighted average importance sampling and defensive mixture distributions," *Technometrics*, vol. 37, no. 2, pp. 185–192, 1995.

[12] A. Owen, *Monte Carlo Theory, Methods and Examples*, 2013, Available at http://statweb.stanford.edu/ owen/mc/.

[13] G. Celeux and G. Govaert, "A classification EM algorithm for clustering and two stochastic versions," *Computational Statistics and Data Analysis*, vol. 14, no. 3, pp. 315–332, 1992.

[14] Anat Levin, Boaz Nadler, Fredo Durand, and William T Freeman, "Patch complexity, finite pixel correlations and optimal denoising," in *European Conference on Computer Vision*. Springer, 2012, pp. 73–86.

[15] A. Levin and B. Nadler, "Natural image denoising: Optimality and inherent bounds," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2833–2840.

[16] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[17] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2233–2246, 2012.

[18] C. Deledalle, F. Tupin, and L. Denis, "Poisson nl means: Unsupervised non local means for poisson noise," in *Image processing (ICIP), 2010 17th IEEE int. conf. on*. IEEE, 2010, pp. 801–804.

[19] A. Rond, R. Giryes, and M. Elad, "Poisson inverse problems by the plug-and-play scheme," *Journal of Visual Communication and Image Representation*, vol. 41, pp. 96–108, 2016.